

# 3 0420  
04/02/01

OIPE

## RAW SEQUENCE LISTING

DATE: 03/26/2001

PATENT APPLICATION: US/09/714,936

TIME: 11:51:33

Input Set : D:\PT\_FL.797.111000

Output Set: N:\CRF3\03262001\I714936.raw

ENTERED  
P.S

6 <110> APPLICANT: Tang, Y. Tom  
7 Zhou, Ping  
8 Goodrich, Ryle  
9 Liu, Chenghua  
10 Asundi, Vinod  
11 Ren, Feiyan  
12 Zhang, Jie  
13 Zhao, Qing A.  
14 Xue, Aidong J.  
15 Yang, Yonghong  
16 Wehrman, Tom  
17 Drmanac, Radoje T.  
20 <120> TITLE OF INVENTION: Novel Nucleic Acids and  
21 Polypeptides  
25 <130> FILE REFERENCE: 797  
C--> 27 <140> CURRENT APPLICATION NUMBER: US/09/714,936  
28 <141> CURRENT FILING DATE: 2000-11-17  
30 <160> NUMBER OF SEQ ID NOS: 362  
32 <170> SOFTWARE: pt\_FL\_genes Version 2.0  
38 <210> SEQ ID NO: 1  
39 <211> LENGTH: 3539  
40 <212> TYPE: DNA  
41 <213> ORGANISM: Homo sapiens  
43 <220> FEATURE:  
44 <221> NAME/KEY: CDS  
45 <222> LOCATION: (1173)..(3539)  
47 <400> SEQUENCE: 1  
48 aattcaacta gagggcagcc ttgtggatgg ccccgaaagca agcctgatgg aacaggatag 60  
50 aaccaaccat gttgagggca acagactaag tccatttcctg ataccatcac ctcccatttg 120  
52 ccagacagaa cctctggcta caaagctcca gaatggaagc ccaactgcctg agagagctca 180  
54 tccagaagta aatggagaca ccaagtggca ctctttcaaa agttattatg gaataccctg 240  
56 tatgaaggga agccagaata gtcgtgtgag tccctgacttt acacaagaaa gtagagggtg 300  
58 ttccaagtgt ttgcaaaatg gaggaataaa acgcacagtt agtgaacctt ctctctctgg 360  
60 gtccttcag atcaagaaat tgaaacaaga ccaaaaggct aatggagaaa gacgtaactt 420  
62 cggggtaagc caagaaagaa atccagggtga aagcagtcac ccaaatgtct ccgatttgag 480  
64 tgataagaaa gaactctgtga gttctgtagc ccaagaaaat gcagttaaag atttcaccag 540  
66 tttttcaaca cataactgca gtgggcctga aaatccagag cttcagattc tgaatgagca 600  
68 ggaggggaaa agtgctaatt accatgacaa gaacattgta ttacttaaaa acaaggcagt 660  
70 gctaattgctt aatggtgcta cagtttctgc ctcttcctg gaacacacac atggtgaact 720  
72 cctggaaaaa acactgtctc aatattatcc agattgtgtt tccattgcgg tgcagaaaac 780  
74 cacatctcac ataaatgcc aataacagtc ggctactaat gagttgtcct gtgagatcac 840  
76 tcacccatcg catacctcag ggcagatcaa ttccgcacag acctctaact ctgagctgcc 900  
78 tccaaagcca gctgcagtgg tgagtgaggg ctgtgatgct gatgatgctg ataatgccag 960  
80 taaactagct gcaatgctaa atacctgttc ctttcagaaa ccagaacaat tatcaacaac 1020  
82 attaatacgt ttttgagata tgcccattct ctgcagaaaa taacatccag ggaaccacaa 1080  
84 agctagcgtc ttgtgaaaga ttctgttcag gttccagcag caatttgcaa gtcctctggtg 1140  
86 gcagctctga acggtattta aaacaaaatg aa atg aat ggt gct tac ttc aag 1193

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/714,936

DATE: 03/26/2001  
 TIME: 11:51:33

Input Set : D:\PT\_FL.797.111000  
 Output Set: N:\CRF3\03262001\I714936.raw

		Met	Asn	Gly	Ala	Tyr	Phe	Lys	
87		1					5		
88									
90	caa agc tca gtg ttc act aag gat tcc ttt tct gcc act acc aca cca								1241
91	Gln Ser Ser Val Phe Thr Lys Asp Ser Phe Ser Ala Thr Thr Thr Pro								
92	10 15 20								
94	cca cca cca tca caa ttg ctt ctt tct ccc cct cct cct ctt cca cag								1289
95	Pro Pro Pro Ser Gln Leu Leu Leu Ser Pro Pro Pro Pro Leu Pro Gln								
96	25 30 35								
98	gtt cct cag ctt cct tca gaa gga aaa agc act ctg aat ggt gga gtt								1337
99	Val Pro Gln Leu Pro Ser Glu Gly Lys Ser Thr Leu Asn Gly Gly Val								
100	40 45 50 55								
102	tta gaa gaa cac cac cac tac ccc aac caa agt aac aca aca ctt tta								1385
103	Leu Glu Glu His His His Tyr Pro Asn Gln Ser Asn Thr Thr Leu Leu								
104	60 65 70								
106	agg gaa gtg aaa ata gag ggt aaa cct gag gca cca cct tcc cag agt								1433
107	Arg Glu Val Lys Ile Glu Gly Lys Pro Glu Ala Pro Pro Ser Gln Ser								
108	75 80 85								
110	cct aat cca tct aca cat gta tgc agc cct tct ccg atg ctt tct gaa								1481
111	Pro Asn Pro Ser Thr His Val Cys Ser Pro Ser Pro Met Leu Ser Glu								
112	90 95 100								
114	agg cct cag aat aat tgt gtg aac agg aat gac ata cag act gca ggg								1529
115	Arg Pro Gln Asn Asn Cys Val Asn Arg Asn Asp Ile Gln Thr Ala Gly								
116	105 110 115								
118	aca atg act gtt cca ttg tgt tct gag aaa aca aga cca atg tca gaa								1577
119	Thr Met Thr Val Pro Leu Cys Ser Glu Lys Thr Arg Pro Met Ser Glu								
120	120 125 130 135								
122	cac ctc aag cat aac cca cca att ttt ggt agc agt gga gag cta cag								1625
123	His Leu Lys His Asn Pro Pro Ile Phe Gly Ser Ser Gly Glu Leu Gln								
124	140 145 150								
126	gac aac tgc cag cag ttg atg aga aac aaa gag caa gag att ctg aag								1673
127	Asp Asn Cys Gln Gln Leu Met Arg Asn Lys Glu Gln Glu Ile Leu Lys								
128	155 160 165								
130	ggg cga gac aag gag caa aca cga gat ctt gtg ccc cca aca cag cac								1721
131	Gly Arg Asp Lys Glu Gln Thr Arg Asp Leu Val Pro Pro Thr Gln His								
132	170 175 180								
134	tat ctg aaa cca gga tgg att gaa ttg aag gcc cct cgt ttt cac caa								1769
135	Tyr Leu Lys Pro Gly Trp Ile Glu Leu Lys Ala Pro Arg Phe His Gln								
136	185 190 195								
138	gcg gaa tcc cat cta aaa cgt aat gag gca tca ctg cca tca att ctt								1817
139	Ala Glu Ser His Leu Lys Arg Asn Glu Ala Ser Leu Pro Ser Ile Leu								
140	200 205 210 215								
142	cag tat caa ccc aat ctc tcc aat caa atg acc tcc aaa caa tac act								1865
143	Gln Tyr Gln Pro Asn Leu Ser Asn Gln Met Thr Ser Lys Gln Tyr Thr								
144	220 225 230								
146	gga aat tcc aac atg cct ggg ggg ctc cca agg caa gct tac acc cag								1913
147	Gly Asn Ser Asn Met Pro Gly Gly Leu Pro Arg Gln Ala Tyr Thr Gln								
148	235 240 245								
150	aaa aca aca cag ctg gag cac aag tca caa atg tac caa gtt gaa atg								1961
151	Lys Thr Thr Gln Leu Glu His Lys Ser Gln Met Tyr Gln Val Glu Met								

DATE: 03/26/2001

TIME: 11:51:33

Output Set: N:\CRF3\03262001\I714936.raw

## RAW SEQUENCE LISTING

DATE: 03/26/2001

PATENT APPLICATION: US/09/714,936

TIME: 11:51:33

Input Set : D:\PT\_FL.797.111000

Output Set: N:\CRF3\03262001\I714936.raw

218	gtt cta cag gga tat aaa aat aga aac caa gat atg tct ggt caa caa	2777
219	Val Leu Gln Gly Tyr Lys Asn Arg Asn Gln Asp Met Ser Gly Gln Gln	
220	520 525 530 535	
222	gct gcg caa ctt gct cag caa agg tac ttg ata cat aac cat gca aat	2825
223	Ala Ala Gln Leu Ala Gln Gln Arg Tyr Leu Ile His Asn His Ala Asn	
224	540 545 550	
226	gtt ttt cct gtg cct gac cag gga gga agt cac act cag acc cct ccc	2873
227	Val Phe Pro Val Pro Asp Gln Gly Gly Ser His Thr Gln Thr Pro Pro	
228	555 560 565	
230	cag aag gac act caa aag cat gct gct cta agg tgg cat ctc tta cag	2921
231	Gln Lys Asp Thr Gln Lys His Ala Ala Leu Arg Trp His Leu Leu Gln	
232	570 575 580	
234	aag caa gaa cag cag caa aca cag caa ccc caa act gag tct tgc cat	2969
235	Lys Gln Glu Gln Gln Gln Thr Gln Gln Pro Gln Thr Glu Ser Cys His	
236	585 590 595	
238	agt cag atg cac agg cca att aag gtg gaa cct gga tgc aag cca cat	3017
239	Ser Gln Met His Arg Pro Ile Lys Val Glu Pro Gly Cys Lys Pro His	
240	600 605 610 615	
242	gcc tgt atg cac aca gca cca cca gaa aac aaa aca tgg aaa aag gta	3065
243	Ala Cys Met His Thr Ala Pro Pro Glu Asn Lys Thr Trp Lys Lys Val	
244	620 625 630	
246	act aag caa gag aat cca cct gca agc tgt gat aat gtg cag caa aag	3113
247	Thr Lys Gln Glu Asn Pro Pro Ala Ser Cys Asp Asn Val Gln Gln Lys	
248	635 640 645	
250	agc atc att gag acc atg gag cag cat ctg aag cag ttt cac gcc aag	3161
251	Ser Ile Ile Glu Thr Met Glu Gln His Leu Lys Gln Phe His Ala Lys	
252	650 655 660	
254	tcg tta ttt gac cat aag gct ctt act ctc aaa tca cag aag caa gta	3209
255	Ser Leu Phe Asp His Lys Ala Leu Thr Leu Lys Ser Gln Lys Gln Val	
256	665 670 675	
258	aaa gtt gaa atg tca ggg cca gtc aca gtt ttg act aga caa acc act	3257
259	Lys Val Glu Met Ser Gly Pro Val Thr Val Leu Thr Arg Gln Thr Thr	
260	680 685 690 695	
262	gct gca gaa ctt gat agc cac acc cca gct tta gag cag caa aca act	3305
263	Ala Ala Glu Leu Asp Ser His Thr Pro Ala Leu Glu Gln Gln Thr Thr	
264	700 705 710	
266	tct tca gaa aag aca cca acc aaa aga aca gct gct tct gtt ctc aat	3353
267	Ser Ser Glu Lys Thr Pro Thr Lys Arg Thr Ala Ala Ser Val Leu Asn	
268	715 720 725	
270	aat ttt ata gag tca cct tcc aaa tta cta gat act cct ata aaa aat	3401
271	Asn Phe Ile Glu Ser Pro Ser Lys Leu Leu Asp Thr Pro Ile Lys Asn	
272	730 735 740	
274	tta ttg gat aca cct gtc aag act caa tat gat ttc cca tct tgc aga	3449
275	Leu Leu Asp Thr Pro Val Lys Thr Gln Tyr Asp Phe Pro Ser Cys Arg	
276	745 750 755	
278	tgt gta gat cct gta aaa ttt gaa tgt atc tgt ttt aga tca att cgc	3497
279	Cys Val Asp Pro Val Lys Phe Glu Cys Ile Cys Phe Arg Ser Ile Arg	
280	760 765 770 775	
282	cta ttt agc tct ttg tat att atc tcc tgg aga gac agc tag	3539

## RAW SEQUENCE LISTING

DATE: 03/26/2001

PATENT APPLICATION: US/09/714,936

TIME: 11:51:33

Input Set : D:\PT\_FL.797.111000

Output Set: N:\CRF3\03262001\I714936.raw

```

283  Leu Phe Ser Ser Leu Tyr Ile Ile Ser Trp Arg Asp Ser  *
284              780              785
290 <210> SEQ ID NO: 2
291 <211> LENGTH: 1342
292 <212> TYPE: DNA
293 <213> ORGANISM: Homo sapiens
295 <220> FEATURE:
296 <221> NAME/KEY: CDS
297 <222> LOCATION: (25)..(894)
299 <400> SEQUENCE: 2
300  cgggccggcg cgggagcaag cggc      atg gcg ttc cgg cag gcg ctg cag ctg      51
301              Met Ala Phe Arg Gln Ala Leu Gln Leu
302              1              5
304  gcg gcc tgc ggg ctg gcc ggg ggc tcg gcc gcc gtg ctc ttc tcg gcc      99
305  Ala Ala Cys Gly Leu Ala Gly Gly Ser Ala Ala Val Leu Phe Ser Ala
306    10              15              20              25
308  gtg gcg gta ggg aag ccg cgc gca ggc ggg gac gcg gag cca cgc ccg      147
309  Val Ala Val Gly Lys Pro Arg Ala Gly Gly Asp Ala Glu Pro Arg Pro
310              30              35              40
312  gct gag ccg ccg gcc tgg gcg ggg ggc cgc cgg ccg gcc ccc ggt gtc      195
313  Ala Glu Pro Pro Ala Trp Ala Gly Gly Ala Arg Pro Gly Pro Gly Val
314              45              50              55
316  tgg gac ccc aac tgg gac agg cga gaa cca ctg tct ctg atc aac gtg      243
317  Trp Asp Pro Asn Trp Asp Arg Arg Glu Pro Leu Ser Leu Ile Asn Val
318              60              65              70
320  cgg aag agg aac gtg gaa tct ggg gaa gaa gag ctg gcg tcc aag ctg      291
321  Arg Lys Arg Asn Val Glu Ser Gly Glu Glu Glu Leu Ala Ser Lys Leu
322              75              80              85
324  gac cac tac aaa gcc aag gcc acg cgg cac atc ttc ctc atc agg cat      339
325  Asp His Tyr Lys Ala Lys Ala Thr Arg His Ile Phe Leu Ile Arg His
326    90              95              100              105
328  tcc cag tac cac gtg gat ggc tcc ctg gag aag gac cgc act ctg acc      387
329  Ser Gln Tyr His Val Asp Gly Ser Leu Glu Lys Asp Arg Thr Leu Thr
330              110              115              120
332  ccg ctg ggt cgg gag cag gct gaa ctc act ggg ctc cgc ctg gca agc      435
333  Pro Leu Gly Arg Glu Gln Ala Glu Leu Thr Gly Leu Arg Leu Ala Ser
334              125              130              135
336  ttg ggg ttg aag ttt aat aaa atc gtc cat tcg tct atg acg cgc gcc      483
337  Leu Gly Leu Lys Phe Asn Lys Ile Val His Ser Ser Met Thr Arg Ala
338              140              145              150
340  ata gag acc acc gat atc atc agc cgg cac ctg cca ggc gtc tgc aaa      531
341  Ile Glu Thr Thr Asp Ile Ile Ser Arg His Leu Pro Gly Val Cys Lys
342              155              160              165
344  gtc agc aca gat ctg ctg cgg gaa ggc gcc ccc atc gag cca gac ccg      579
345  Val Ser Thr Asp Leu Leu Arg Glu Gly Ala Pro Ile Glu Pro Asp Pro
346    170              175              180              185
348  ccc gtg tct cat tgg aag ccg gaa gct gtg cag tat tac gaa gac gga      627
349  Pro Val Ser His Trp Lys Pro Glu Ala Val Gln Tyr Tyr Glu Asp Gly
350              190              195              200

```

FII:

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

## VERIFICATION SUMMARY

PATENT APPLICATION: US/09/714,936

DATE: 03/26/2001

TIME: 11:51:34

Input Set : D:\PT\_FL.797.111000

Output Set: N:\CRF3\03262001\I714936.raw

L:27 M:270 C: Current Application Number differs, Replaced Current Application Number  
L:374 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:2  
L:531 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3  
L:934 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:6  
L:1195 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:1625 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:9  
L:2054 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:12  
L:2961 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:20  
L:4323 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:30  
L:5723 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:37  
L:6638 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:42  
L:6988 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45  
L:6990 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45  
L:6992 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45  
L:6994 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45  
L:6996 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45  
L:7550 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:49  
L:7797 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:50  
L:7937 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:52  
L:8080 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:53  
L:8708 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:57  
L:9741 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:64  
L:9940 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:65  
L:10070 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:67  
L:10249 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:70  
L:10518 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:10520 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:10522 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73  
L:10595 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:74  
L:11158 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:78  
L:11742 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:80  
L:11970 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:82  
L:12662 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:88  
L:12983 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:89  
L:13007 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:90  
L:13045 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:90  
L:13482 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:95  
L:15582 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:111  
L:15925 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:115  
L:15980 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:116  
L:16328 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:118  
L:16823 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:121  
L:16951 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:122  
L:16979 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:123  
L:17170 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:124  
L:17793 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:129  
L:18475 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:138  
L:18541 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:139

VERIFICATION SUMMARY  
PATENT APPLICATION: US/09/714,936

DATE: 03/26/2001  
TIME: 11:51:34

Input Set : D:\PT\_FL.797.111000  
Output Set: N:\CRF3\03262001\I714936.raw

L:18838 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:141  
L:18982 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:142  
L:19191 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:144  
L:19230 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:145  
L:19387 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:147  
L:19954 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:152  
L:20161 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:153  
L:20268 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:154  
L:20438 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:156  
L:20545 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:157  
L:20870 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:160  
L:21036 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:162  
L:21109 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:163  
L:21230 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:164  
L:21390 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:166  
L:21962 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:168  
L:22720 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:176  
L:22892 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:178  
L:26259 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:207  
L:26857 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:212  
L:30137 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:239  
L:32576 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:260  
L:33279 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:269  
L:33708 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:275  
L:34295 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:281  
L:38209 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:326  
L:38553 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:331  
L:38914 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:335  
L:38916 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:335  
L:38918 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:335  
L:40136 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:349  
L:40710 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:356  
L:40714 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:356